图书情报工作
第 63 卷 第 1 期 2019 年 1 月

研究数据管理服务框架研究与构建:以北京大学为例

崔海媛 罗鹏程 李国俊 朱玲
北京大学图书馆 北京 100871

摘要:[**目的/意义**]通过数据驱动研究的海内外环境扫描,研究数据管理服务的国内外进展,提出研究数据管理服务框架和服务研究生命周期的建议,并以北京大学图书馆的实践为例,为高校与科研机构提供研究数据管理服务、推动数据驱动研究提供研究与实践参考经验。[**方法/过程**]调研国内外数据驱动研究现状和趋势,分析研究数据管理服务国内外发展情况,指出研究数据管理服务的重要性。通过对用户调查和 QS 排名前 50 名的大学服务的调研,总结服务需求。根据数据驱动研究范式特点和研究者需求,依据国家、国际组织、资助机构和科研机构的数据政策要求,基于研究数据生命周期需要服务于研究生命周期,提出研究数据管理服务框架设计。介绍北京大学研究数据管理服务框架体系建设经验,为研究数据管理服务框架体系建设提供建议。[**结果/结论**]基于需求调研设计支持学术交流全过程的研究数据管理服务框架,研究数据管理服务应包括:数据管理计划、数据服务、数据管理、数据素养、长期保存和数据政策等。在该框架下北京大学图书馆从多方面开展研究数据服务实践,为高校与科研机构建设研究数据管理服务提供有价值的参考经验。

关键词:数据驱动研究 研究数据管理 数据服务框架 数据服务 数据政策
分类号:G251
DOI:10.13266/j.issn.0252-3116.2019.01.012

1 引言

数据与普通民众的生活、工作、学习和研究有何关系?数据的影响到底有多大?2018 年 3 月曝光的政治 AI 公司剑桥分析(Cambridge Analytica, AI)通过大数据分析影响美国大选事件,让普通人感受到大数据已经与个人生活密不可分。剑桥分析公司首席执行官 A. Nix 说:"特朗普的每一条选举信息都是数据驱动的"[1]。2018 年 3 月 27 日,英国剑桥分析公司前员工克里斯托弗.韦利,在英国议会接受质询提到:如果没有我所认为的作弊(行为),"脱欧"公投结果很可能会截然不同[2]。剑桥分析公司事件让民众了解到:大数据和人工智能环境下,数据已经影响所有人。

研究表明,数据驱动已经是全世界、全领域、全行业发展态势。

经济领域,2017 年以来,全球市场高度关注人工智能和大数据等数据驱动经济发展。有专家预测,至 2022 年,全球大数据市场规模将达到 800 亿美元,年均实现 15.37% 的增长。2017 年全球的数据总量为 21.6ZB,增长速度在每年 40% 左右,预计到 2020 年全球的数据总量将达到 40ZB[3]。

国家战略层面,美国、欧盟、日本、韩国和中国等国纷纷将大数据发展上升为国家战略,积极推进数据政策,希望利用大数据技术实现研究和产业领域突破[4]。美国加大在科研教学、环境保护、工程技术、国土安全、生物医药等领域的投入,研发计划涉及自然科学基金委、国家卫生研究院、国防部、能源部、国防部高级研究局、地质勘探局等多个联邦部门和机构。欧盟实施研究数据价值链战略,资助"大数据"和"开放数据"领域的研究和创新活动,推动实施开放数据政策,促进公共资助科研实验成果和数据的使用及再利用。日本推出以开放大数据为核心的国家战略,目标是建成具有"世界最高水准的广泛运用信息产业技术的社会"。韩国提出"培育 1 000 家大数据、云计算系统相关企业"的大数据发展计划,出台《第五次国家信息化基本计划(2013-2017)》等多项大数据发展战略[3]。中国在十九大报告中,首次提出信息化领域 3 个全新概念:网络强国、数字中国、智慧社会[5]。

黄铁军提出,强人工智能时代即将来临,2030 年,

作者简介:崔海媛(ORCID:0000-0001-5541-7100),副研究馆员,硕士,E-mail:cuihy@lib.pku.edu.cn;罗鹏程(ORCID:0000-0001-9598-0715),馆员,硕士;李国俊(ORCID:0000-0002-4260-0560),馆员,硕士;朱玲(ORCID:0000-0003-2881-8050),馆员,硕士。

收稿日期:2018-10-16 修回日期:2018-11-21 本文起止页码:81-90 本文责任编辑:王传清

结构仿脑、功能类脑、性能超脑的神经计算机就可能问世[6]。可以预见,人工智能的发展,会大大增加数据的复杂度。如何管理和使用数据,将成为人脑的挑战。大数据与人工智能快速发展,数据已经成为科学研究、学术交流、技术创新、工业发展、社会媒体与社交网络、教育教学、文化生活与经济发展的基础。新的学术交流环境,图书馆服务内容和方式均面临挑战,急需转型。

本文将调研国内外数据驱动研究现状和数据服务需求,分析研究数据管理(research data management,RDM)服务国内外政策和发展情况,提出研究数据管理服务框架,并介绍北京大学研究数据管理服务的研究成果、框架设计、构建经验与服务效果,总结研究数据管理服务框架体系建设经验,为推动研究数据管理服务提供建议。

2 研究数据管理服务研究进展和趋势

2.1 数据驱动研究服务需求

高等教育变革,数据科学和数据驱动研究成为全球高校发展重点,需要数据管理服务支撑教学科研。在studyportals网站统计中,美国有104所高校设置"Data Science & Big Data"本科专业。2016年2月,北京大学、对外经济贸易大学、中南大学首批成功申请到"数据科学与大数据技术"本科专业。2017年3月,第二批申请中共有32所高校获批。2018年3月,有248所学校获批,是过去两次审批通过额度的近8倍[7]。数据科学已成为当下高校的热门专业领域,需要教学科研服务支撑。

数据驱动研究已经成为全学科研究和发展态势,数据服务需求日益迫切。2017年11月24日至2018年4月3日,由北京大学图书馆、北京大学信息管理系和海南省南海大数据应用研究院发起与组织,多家机构合作,成功举办了"首届全国高校数据驱动创新研究大赛"(以下简称"大赛")。大赛吸引来自全国169所高校的1 892名同学(共593支队伍)参赛报名,报名队伍分布于56个一级学科[8]。大赛选手的广泛参与和参赛作品分布于全学科,说明数据驱动研究已经成为全学科研究和发展态势,用户对数据的需求已经逐渐超越对文献的需求,使用数据资源,基于数据进行研究已经是趋势。

数据日益重要,重视与发展研究数据的管理与服务已成为国家层面与科研机构未来提升学术竞争力和影响力的重要发展战略之一。2018年4月2日,国务院办公厅印发了《科学数据管理办法》[9],以研究数据全生命周期为中心,加强和规范研究数据的管理,推动开放共享。在这样的形势下,建设一流的研究数据管理服务框架体系,提供研究数据汇集、管理与服务,满足用户数据驱动研究需求,成为一流高校的必然选择。

2.2 研究进展和趋势

2.2.1 数据政策推动服务发展 国际上,数据政策陆续发布,推动了全球研究数据管理和开放服务的发展。2011年,美国自然科学基金委要求项目申请书必须包含"数据管理计划",美国多个资助机构数据管理政策陆续发布[10]。2011年,英国"国家数据监管中心(National Digital Curation Center, DCC)"发布《制定数据管理与共享计划》指南[11]。2012年7月,英国皇家学会(Royal Society)发布《科学是开放事业(Science as an open enterprise)》报告,确立共享和开放科学数据的基本原则、机遇与面临的挑战[12]。2013年6月,"八国集团"首脑在北爱尔兰"G8 Summit峰会"上签署《开放数据宪章》,该宪章中提出5点要求:①使开放数据成为规则;② 注重数据质量和数量;③让所有人都可以使用;④改善数据发布的管理流程;⑤发布数据为创新服务[13]。2014年,芬兰Academy of Finland提倡研究者开放共享研究成果和数据[14]。欧盟2020计划规定,2017年开始全面实施科研数据开放制度,推动"开放科学"战略[15]。

2015年8月31日,中国国务院以国发〔2015〕50号印发《促进大数据发展行动纲要》,发展科学大数据[16]。积极推动由国家公共财政支持的公益性科研活动获取和产生的科学数据逐步开放共享,构建科学大数据国家重大基础设施,实现对国家重要科技数据的权威汇集、长期保存、集成管理和全面共享。2014-2015年,国家自然科学基金委发布开放获取政策和细则[17],并发布资助成果基础知识库[18],正式开启了中国资助机构的开放获取行动,但目前仍没有发布开放数据政策和开放数据平台。

全球数据仓储快速发展,截至2018年10月,根据国际研究数据仓储注册组织(Registry of Research Data Repositories, Re3data)统计,全世界至少有2 224多个研究数据仓储[19]。截至2018年10月,在科学数据标识符注册中心(DataCite)中已注册数据集(Dataset)类型DOI为5 123 830个[20]。

2.2.2 服务框架研究综述 需求催生服务,研究数据管理成为研究型高校和图书馆的重要服务之一,相关研究与实践发展迅速。目前国内外关于研究数据管理服务框架的相关研究,主要集中在服务调研、服务内容

框架、实践案例、数据素养教育等方面。

服务调研方面，2014 年，DCC 发布了对来自英国 67 个研究机构共 87 位管理人员的调查结果，调查显示各研究机构对 RDM 的资金、人员和基础设施的投入并不均衡[21]。C. Tenopir 等在 2016 年 2 月通过邮件调查了 119 个欧洲大学图书馆研究数据管理服务的开展情况[22]，结果显示，90% 以上图书馆都已经开展或计划开展数据政策、数据管理计划（data management plan，DMP）、数据素养培训等服务，但仅有 31.9% 的图书馆提供了研究项目中的数据服务。A. M. Cox 等在 2017 年，对 170 个来自澳大利亚、加拿大、德国、爱尔兰、苏格兰、新西兰和英国的图书馆进行调查，结果显示，与高度认可数据政策与开展服务的意愿相比，各机构整体缺乏高质量、高水平的 RDM 服务[23]。A. Yoon 等在 2015 年对 185 个美国研究型大学图书馆网站进行调查分析，发现提供的研究数据管理服务主要包括：咨询帮助（help/ask librarians，79.5%）、数据管理（data deposit，60%）、研究数据管理计划（data management planning，41.1%）、数据咨询（data consultation，38.4%）、数据出版和共享（data publishing and sharing，27.6%）、数据方法（data methodology，15.7%）[24]。

服务内容框架方面，2017 年 3 月，OCLC 发布“研究数据管理现实”系列报告，研究总结了 RDM 服务的三大领域：教育类（数据意识、数据长期保存概念、数据政策、数据管理方法、数据管理工具软件、数据资源）、解决方案类（数据馆员面向研究者研究过程的个性化需求服务）和长期保存类（研究生命周期中的数据管理服务、标识符、权限管理、元数据管理、版本管理、长期保存）[25]。澳大利亚国际数据服务中心（The Australian National Data Service，ANDS）提出数据管理能力的 5 个要素：政策和流程、信息架构、服务能力、元数据管理和研究数据管理[26]。英国数据管理中心（The Digital Curation Centre，DCC）根据研究生命周期，建议研究数据管理支持服务框架应涵盖：政策和规划、可持续性发展、研究数据管理计划、元数据管理、数据存储、数据存档策略、数据管理、培训指南等（见图 1）[27]。

崔宇红参考信息通信技术促进发展项目（Information and Communication Technology for Development，ICT4D）提出 RDM4D 模型立方体，并通过对 20 所英国大学研究数据管理服务内容的分析和归纳，依据研究过程的 3 个不同阶段将 21 个关键实践聚类到 4 个关键过程域中，即：基金项目申请前创建研究数据管理计划；基金项目进行中组织、描述和存储研究数据；基



图 1　DCC 研究数据管理支持服务框架

金项目完成后对研究数据的长期保存和访问；数据发布和共享[28]。

实践案例方面，国内的复旦大学、北京大学和武汉大学已经率先提供研究数据管理服务，但均无数据政策发布[29－31]。国际上，因国家和资助机构政策要求，服务已经广泛开展，相关研究案例很多，内容集中在分享各机构的服务实践经验[32－34]。以昆士兰大学为例，在用户调研基础上，提供支持研究生命周期的数据服务内容包括：①项目规划和准备阶段（资金和道德批准、知识产权和法律问题）；②项目实施阶段（数据收集、处理和分析）；③存档、发布和传播阶段（传播研究成果和数据存档以便共享和再利用）[35]。

文献研究发现，国内目前对研究数据管理服务相关研究和实践相对不足，研究成果不多，研究集中在国际政策、服务和需求调研方面，数据平台建设和服务开展进展缓慢。国际上，研究内容以调研和实践案例介绍为主，对服务框架的研究不深入，缺乏对框架内容的具体研究、构建和建议。

3　支持学术交流全过程的研究数据管理服务框架设计

3.1　需求调研

2014 年，笔者团队对校内教师开展问卷调查，获取了 50 份教师问卷的调查结果。随后对校内 26 名研究与数据密切相关的教师进行一对一需求访谈。调研发现：高达 87.5% 的受访者愿意在一定条件下开放研

究数据，开放的最大动机是通过开放数据提升数据价值，提高成果引用率与增强数据曝光度。开放的最大疑虑是他人利用数据抢先发表成果。开放条件中，受访者最需要“审核使用者身份和数据使用用途”功能。

2018 年，为了找出研究数据管理的核心工具和服务，多伦多大学两位图书馆员从多伦多大学招募了 28 名教师，通过焦点小组法研究发现了研究者关心的 4 类 RDM 服务：①支持研究过程的基础设施；②数据安全；③数据管理技能；④发布数据的版权保护[36]。

3.2　RDM 服务调研

笔者对 2018 年 QS 世界大学排名(Quacquarelli Symonds World University Rankings)前 50 所大学(美国 18 所、加拿大 2 所、英国 9 所、澳大利亚 5 所、瑞士 2 所、中国大陆和香港 7 所、法国 1 所、新加坡 2 所、日本 2 所，韩国 2 所)进行调研，了解是否提供研究数据管理服务，提供哪些服务，梳理国内外研究型大学研究数据管理服务内容。研究发现提供的研究数据管理服务内容主要包括：数据管理计划、数据服务、数据素养、数据管理、长期保存和数据政策等几部分。调研同时发现：政策支持下，前 50 所大学除了亚洲国家大学外，基本都有研究数据管理服务，亚洲国家大学提供研究数据管理服务的较少；在服务内容上，开展较好的服务是数据管理计划和数据服务，数据管理和长期保存等相关服务有待提供与加强。

表 1　2018 年 QS 世界大学排名前 50 所大学研究数据管理服务内容调研结果

大学名称	数据管理计划	数据服务				数据素养					数据管理			长期保存				数据政策
	DMP指南	数据需求咨询	数据共享	数据引用	数据版权	数据生命周期介绍	数据政策指南	DMPTool培训	工具软件培训	数据仓储指南	元数据描述	文件格式	文件命名	数据存储	机构的数据仓储	数据安全与备份	软件代码保存	机构的数据政策
麻省理工学院	√	√					√	√	√		√	√	√	√				
斯坦福大学	√	√	√	√	√			√	√	√	√	√	√	√	√	√		
哈佛大学	√	√		√			√	√		√		√	√	√	√			
加州理工学院	√	√	√					√			√	√	√	√	√			
剑桥大学	√	√	√	√			√		√		√	√	√		√			√
牛津大学	√	√	√	√	√		√		√		√		√	√	√	√		√
伦敦大学学院	√		√	√	√		√					√		√	√		√	√
帝国理工大学	√		√							√	√	√	√			√		√
芝加哥大学	√		√					√		√	√	√	√			√		
苏黎世联邦理工学院	√	√	√						√		√	√	√		√	√		
南洋理工大学	√		√		√		√			√					√			√
洛桑联邦理工学院	√					√				√	√	√				√		
普林斯顿大学	√		√	√	√						√	√	√		√	√		
康奈尔大学	√	√	√	√	√		√				√	√	√	√	√	√		
新加坡国立大学	√		√	√	√		√			√	√	√	√		√			√
耶鲁大学	√	√		√					√	√	√							
约翰霍普金斯大学	√	√	√	√	√		√		√	√	√			√	√	√	√	√
哥伦比亚大学	√	√	√	√	√	√	√			√	√	√	√	√	√	√		√
宾夕法尼亚大学	√		√	√			√	√		√	√	√	√			√		
澳大利亚国立大学	√		√	√	√					√	√	√	√		√	√		
杜克大学	√			√							√	√	√		√	√		
密歇根大学	√		√	√							√				√	√		
爱丁堡大学	√	√					√		√						√			√
伦敦国王学院	√	√	√	√	√	√	√		√		√	√	√	√	√	√		
清华大学(无)																		
香港大学				√	√					√	√	√	√			√		√
加州伯克利大学	√	√	√	√	√						√			√	√	√		
东京大学(无)																		
西北大学	√	√		√			√			√	√				√			
香港科技大学	√			√		√				√	√				√			
多伦多大学	√		√	√	√		√				√	√	√		√	√		
麦吉尔大学	√		√	√														
加州大学洛杉矶分校	√			√	√					√	√	√	√	√		√		
曼彻斯特大学	√		√	√			√				√	√	√	√	√	√		
伦敦政治经济学院	√										√	√	√			√		√
京都大学(无)																		

（续表1）

大学名称	数据管理计划	数据服务				数据素养					数据管理			长期保存				数据政策
	DMP指南	数据需求咨询	数据共享	数据引用	数据版权	数据生命周期介绍	数据政策指南	DMPTool培训	工具软件培训	数据仓储指南	元数据描述	文件格式	文件命名	数据存储	机构的数据仓储	数据安全与备份	软件代码保存	机构的数据政策
首尔大学(无)																		
加州大学圣迭戈分校	√										√	√	√	√		√		
北京大学		√	√	√	√				√		√	√	√	√	√	√	√	
复旦大学															√			
韩国科学技术院(无)																		
墨尔本大学	√	√					√	√	√									√
巴黎高等师范学校(无)																		
布里斯托大学	√		√											√				√
新南威尔士大学	√		√	√	√						√				√			
香港中文大学(无)																		
昆士兰大学	√		√			√	√		√					√	√			
卡耐基梅隆大学	√		√	√	√	√	√				√	√	√	√				
香港城市大学(无)																		
悉尼大学				√	√				√		√	√	√	√				√

3.3 框架设计

调研发现，数据驱动研究者需要的数据服务主要包括：①满足国际资助机构申请要求的服务；②支持教学、科研和数据发布的数据服务；③研究过程所需的数据管理基础设施和服务；④数据驱动所需的数据技能和知识；⑤数据长期保存服务；⑥数据政策、数据素养和知识产权服务。

根据数据驱动研究范式特点和研究者需求，考虑到研究数据生命周期与研究生命周期密切相关（见图2），并依据国家、国际组织、资助机构和科研机构的数据政策要求，结合实践经验，笔者给出研究数据管理服务框架设计内容（见表2），研究数据管理服务应包括：数据管理计划、数据服务、数据管理、数据素养、长期保存和数据政策6个部分。每一部分应包括的服务内容见表3。各机构可以参考框架内容，并根据实际需求设计服务体系，开展RDM服务。



图2 研究数据生命周期和研究生命周期关系

在服务框架下，为支持新一代学术交流生态系统[37]对数据服务的需求（见图3）和高校研究支持服务，建议提供支持学术交流生态系统各研究周期完整数据需求的服务（见图4）。在科研准备阶段，需要了解把握研究领域前沿方向，跟踪研究热点，为项目申请做准备。以数据管理计划、数据服务和数据素养培训为核心，满足研究者需求。在搜索发现阶段，需要获取完整文献和数据资源，以建设海量数字资源导航和开

表 2　研究数据管理服务框架

数据管理计划	数据服务	数据管理	数据素养	长期保存	数据政策
DMP 指南	RDM 介绍（概念、意义、生命周期）	RDM 工具软件服务	DMP 工具使用	框架体系	RDM 政策制定与发布
DMP 咨询	国际发展动态	RDM 私有云服务	数据素养培训	元数据标准	知识产权培训
	开放数据平台	项目过程管理支持	工具软件使用	存储保存	政策培训（国家、资助机构、机构、期刊）
	研究数据支持		研究培训	安全备份	政策咨询
	教学数据支持		写作培训	源代码保存	
	数据需求咨询		数据资源导航		

表 3　研究数据管理服务内容说明

服务框架	服务内容	服务说明
数据管理计划	DMP 指南	为研究者提供 DMP 写作指南、资助机构 DMP 要求，机构 DMP 政策等服务支持
	DMP 咨询	由了解 DMP 的数据馆员为研究者提供咨询服务
数据服务	RDM 介绍（概念、意义、生命周期）	提供 RMD 概念、意义和生命周期等服务入口（网站、手册等），帮助研究者了解 RDM
	国际发展动态	提供国家、国际组织、资助机构等 RDM 最新发展动态（网站、手册等），提升研究者数据素养，帮助研究者了解最新研究进展
	开放数据平台	高质量研究数据发布与检索服务，为研究者提供数据发布和检索入口。数据学科馆员提供数据整理、分类和归档等服务
	研究数据支持	通过数据馆员、学科馆员、平台和工具软件服务等，为研究者提供研究过程中数据服务支持。
	教学数据支持	以开放数据平台为基础，建设教学数据服务，汇集全学科数据，提供教学支持
	开放数据服务（共享、分享、引用）	以开放数据平台为基础，为研究者提供共享数据、分享传播、引用等开放学术交流服务
	数据需求咨询	设置学科馆员和数据馆员岗位，为研究者提供完整研究生命周期的数据需求咨询服务
数据管理	RDM 工具软件服务	研究过程中的数据管理软件、工具与服务平台，为项目团队提供数据管理支持
	RDM 私有云服务	研究过程中数据管理私有云服务
	项目过程管理支持	研究过程中项目管理平台和服务支持
数据素养	DMP 工具使用	研究管理计划撰写工具使用培训和方法指南
	数据素养培训	开展相关的数据素养培训和咨询，包括文献管理软件的培训和利用咨询，论文写作规范和学术道德规范培养等
	工具软件使用	数据管理工具软件（R、Python 等）使用培训
	研究培训	持续开展数据管理相关资源、软件工具和方法等系列培训，提升用户数据素养
	写作培训	开展论文写作技巧与数据引用、出版合同签署、出版媒介选择、数据版权和发布等服务支持
	数据资源导航	建设数字资源导航，提供数据资源搜索发现服务入口
长期保存	框架体系	建设数字资源长期保存体系规划，保障数据资源长期保存
	元数据标准	制定数据资源元数据标准规范
	存储保存	提供数据存储保存基础设施、平台和服务
	安全备份	制定数据安全备份制度
	源代码保存	源代码保存平台，为研究者提供项目源代码长期保存服务
数据政策	RDM 政策制定与发布	研究制定与推动机构研究数据管理和服务政策，推进开放学术交流
	知识产权培训	数据、成果等知识产权培训与服务，帮助研究者了解知识产权法律法规
	政策培训（国家、资助机构、机构、期刊）	数据政策培训，帮助研究者了解国际、国际组织、资助机构、机构和期刊数据政策，顺利完成项目申请和成果发表
	政策咨询	数据馆员和学科馆员为研究者提供数据政策咨询

放数据服务平台为基础，提供数据发现服务。在研究分析阶段，用户需要项目管理和数据服务支持，以数据管理和长期保存服务提供研究过程需要的数据管理和长期保存所需软件、工具与服务，以数据服务满足研究过程中需要的数据支持服务。在成果写作阶段，以数据开放服务和写作培训为基础，学科馆员和数据馆员服务为主体，为研究者提供论文写作和数据服务。在成果发表阶段，为研究者提供数据知识产权政策咨询服务，帮助研究者顺利发表研究和数据成果，提高学术影响力。在成果推广阶段，以数据服务平台和长期保

存平台，为研究者提供研究数据的发表、存档和传播服务，并支持教学过程，帮助研究者广泛传播学术影响力。在科研评价阶段，提供数据服务和长期保存服务，为完整、精准评价服务提供支持，为研究者总结研究成果、开始新研究提供支持。以完整支持学术交流完整生命周期为目标，设计与构建研究数据管理服务。



图3　新一代学术交流生态系统框架



图4　支持学术交流生态系统各研究周期完整数据需求的服务框架

4　北京大学研究数据管理服务框架设计与服务

4.1　服务框架设计

为适应高校教学和研究变革趋势，顺应大数据时代和数据密集型研究新范式，北京大学图书馆将构建新一代学术交流生态系统，建设数据资源和研究数据管理服务，作为重要发展目标，调整组织机构，重新组织与设计服务，推动数据资源建设和研究数据管理服务。2015 年 12 月 25 日，正式发布开放研究数据服务平台（以下简称“开放数据平台”）、机构知识库、开放出版－期刊网和学者主页 4 个学术成果生态系统服务，为学术交流提供支持，成为国内高校相关建设的引领者。

2015 年，北京大学图书馆调整组织机构，设置研究数据管理服务岗位（见表 4）。在国内外调研和用户调查的基础上，结合北京大学实际需求，设计数据管理服务框架（见表 5），开展平台、服务、培训、政策研究与推进等一系列研究数据管理服务工作，提供研究数据汇集、管理与服务。在组织结构和资源建设上，做以下调整：①原系统部更名为信息化与数据中心，设置数据管理与服务岗位，全面开展研究数据管理平台、工具和服务建设。②组建研究支持中心，设置数据馆员，开展信息素养和数据咨询等服务。③数据资源成为资源建设的重要内容。

表 4　北京大学图书馆研究数据管理服务岗位设置

机构设置	岗位	职责
信息化与数据中心	数据管理与服务组	数据管理与服务 数据政策研究推动 数据管理服务平台建设 长期保存体系建设
研究支持中心	数据馆员	数据素养 数据服务 数据导航
资源建设中心	数据资源建设岗位	数据资源建设

表5 北京大学研究数据管理服务框架

数据服务	数据管理	数据素养	长期保存	数据政策
数据服务主页(建设中)	数据管理工具软件服务(建设中)	数据素养培训	框架体系(建设中)	RDM 政策制定与发布(建设中)
开放数据平台(发布)	数据管理私有云服务(建设中)	工具软件使用	元数据标准	知识产权培训
数据需求咨询	数据管理平台(建设中)	研究培训	存储保存	政策培训(国家、资助机构、机构、期刊)
数据驱动研究大赛		写作培训	安全备份	政策咨询
		数据资源导航	源代码保存(建设中)	

4.2 北京大学研究数据管理服务实践

4.2.1 数据服务 2013年,北京大学开始对研究数据管理进行调研,调研范围包括现有研究数据管理服务案例(如哈佛大学、爱丁堡大学、莫纳什大学、复旦大学等)、数据政策(如美国自然科学基金、英国研究理事会等)、系统平台(如 Dataverse、ICPSR、Data Conservancy、CKAN、Datahub 等)、持久标识符(如 DOI、Handle 等),同时与社科部、科研部、相关院系合作,通过问卷调查和访谈两种形式对教师的研究数据管理需求和现状进行调查。在前期广泛调研的基础上,选择以哈佛大学开源系统 Dataverse 为基础,构建北京大学开放数据服务平台。2015年,图书馆与社科调查中心合作,以国家自然科学基金"北京大学管理科学数据中心"的"数据服务"子项目为契机,共同建设开放数据服务平台。图书馆基于 Dataverse 4.0,增加了双语界面、统一认证、用户审核管理、DOI 注册、数据统计等定制化功能,同时负责系统管理和服务器运维工作;社科调查中心推动数据集的收集、整理和发布。2015年12月,在调研和用户调查基础上,发布基于开源软件 Dataverse 的开放数据平台,为全球用户提供研究数据提交、管理和开放服务。2016年,实现平台数据与机构知识库的成果相互关联,提升数据和成果的曝光度与影响力。2017年,与国际数据仓储 Re3data、国际数据索引平台 Data Citation Index(DCI)实现数据交互、收录与揭示,推动开放数据平台建设成为国际一流研究数据服务平台。截至2018年10月,开放数据平台发布数据空间35个、数据集234个、数据文件1 104个,收录中国家庭追踪调查、中国健康与养老追踪调查、综合型语言知识库、中国历代人物传记资料库等多个高质量、具有国际影响力的数据集。

在平台建设基础上,开始建设数据服务主页,为用户提供完整研究数据服务入口。专业数据馆员提供数据咨询、数据素养培训、数据资源导航等数据咨询服务。OCLC"研究数据管理现实"报告中针对数据咨询特别提出:数据咨询服务必须是馆员提供的面向研究者面对面的咨询服务,而不是邮件、电话和在线咨询服务。

2017年,通过举办大赛,推动全国高校师生利用开放数据进行数据驱动研究的同时,极大地提升了数据平台的影响力。大赛活动期间开放数据平台日均访客数为753人,大赛结束后日均访客数为477人。通过大赛,平台日均访问量增加了近10倍。大赛期间,数据下载量达到过去两年总和的5倍,新增用户数量达到过去两年总和的4倍。2018年,第二届大赛已经在筹备中,通过大赛促进高校数据驱动研究和数据服务,成为北京大学图书馆开展研究数据服务的成功经验。该案例说明,创新服务需要用创新的方法进行推广。

4.2.2 数据管理 如前所述,国际上各机构开展研究数据管理服务,目前多局限于数据素养和数据服务层面,而为研究者提供研究过程中的数据管理服务并不多见,但却是研究者需要的服务。美国明尼苏达大学图书馆在2017年发布了一项关于美国顶尖研究型大学实验室电子笔记(Electronic Laboratory Notebooks, ELNs)应用情况的调查,在35个被调查大学中,有8个大学购买了 ELNs,占比23%。在没有购买的大学中,有14个大学在一些课堂教学或者实验室中使用了 ELNs,占比40%。此外,有7个大学的图书馆提供 ELN 指南或者相关的网页,占比20%。由此可见,美国顶尖研究型大学有超过半数正在应用实验室电子笔记于教学和科研之中,并且一部分图书馆也开始提供实验室电子笔记相关服务[38]。笔者团队已经开始推进数据管理工具软件、私有云和研究过程中的数据管理平台服务工作,以专职数据服务人员、平台和工具软件等多种服务,为研究者提供研究过程中数据管理服务。

4.2.3 数据素养 自2014年开始,数据馆员在北京大学图书馆品牌培训课程"一小时讲座"中,开展数据素养培训,内容包括数据资源、软件工具、数据分析和挖掘工具软件使用和学术规范等培训。2018年开始发布数据资源导航。截至2018年10月,提供了422种数据资源浏览和检索入口[39]。

4.2.4 长期保存 北京大学图书馆将数字资源长期保存体系建设作为重点规划项目,将自建数字加工资源、数据资源等统一考虑,建设长期保存标准、规范与平台,与科

研管理系统、校信息化建设项目融合发展,提供北京大学数字资源长期保存服务。目前已经完成前期调研,正在进行平台构建与开发,计划于2019年推出服务。

4.2.5 数据政策 笔者团队前期完成了北京大学开放获取政策的调研、起草与论证工作,进一步修订后将提交校方。在"一小时讲座"和数据服务主页中,介绍数据政策(国家、资助机构、机构、期刊),帮助研究者了解国际、国际组织、资助机构、机构和期刊的数据政策。2018年10月,北京大学知识产权信息服务中心正式成立,负责建设北京大学知识产权信息资源平台,开展北京大学知识产权培训,为重大科研项目提供科研全周期的知识产权信息咨询服务,参与北京大学产学研协同创新,为知识产权转移转化工作提供协助等。

4.3 未来规划

完善研究数据服务框架,深入服务,建设完整研究数据服务体系,为师生提供高质量的服务,将北京大学建设成为研究数据汇集、管理与服务的世界一流引领者,是图书馆的重要发展目标之一。加强在数据工具、软件、管理与服务等方面研究与应用,推动大数据研究与应用。加快研究数据建设,以数据服务为凝聚力,建设持续汇集国内外高质量研究数据的研究数据平台,提供基于数据的教学与研究的支撑平台。开展数据创新研究、教学与产业孵化活动与比赛,与校外数据资源丰富的政府、企事业单位合作,寻求优势互补、共赢的合作模式,推动数据教学与科研创新。

5 结语

在大数据和人工智能时代,数据驱动研究方法已经被各个学科应用,基于数据进行研究已经是趋势。重视与发展研究数据的管理与服务,已成为世界各国科研机构未来提升学术竞争力和影响力的重要发展战略之一。2018年4月国务院办公厅发布科学数据管理办法后,可以预见,中国资助机构与科研机构的数据政策即将出台,研究数据管理服务也将成为图书馆重要的创新服务内容。图书馆应了解研究数据管理服务框架,规划本机构的服务内容,支持数据驱动研究,为基于数据的教学与研究提供支持,助力中国高等教育双一流建设和科技创新发展。

参考文献:

[1] 钱天培, 邱猛, 龙牧雪, 等. 助特朗普胜选、英国脱欧,深扒Facebook丑闻背后的神秘数据公司[EB/OL]. [2018 - 04 - 15]. http://36kr.com/p/5124932.html.

[2] LANXON N. Whistle-blower says brexit vote could have gone other way[EB/OL]. [2018 - 04 - 15]. https://www.bloomberg.com/news/articles/2018-03-27/whistle-blower-says-brexit-referendum-could-have-gone-other-way.

[3] 2018全球大数据产业将呈七大发展趋势[EB/OL]. [2018 - 04 - 21]. http://www.xinhuanet.com/fortune/2018-02/05/c_1122367340.htm.

[4] 张影强, 张大璐, 梁鹏. 发达国家如何布局大数据战略[J]. 中国经济报告, 2018(1): 87 - 89.

[5] 王晨光. 推进"互联网+政务服务"建设 积极构建数字政府[J]. 北方经济, 2018(Z1): 46 - 48.

[6] 黄铁军. 智能为用, 机器为体[J]. 科学大观园, 2018(2): 1.

[7] 283所高校获批数据科学与大数据专业[EB/OL]. [2018 - 05 - 10]. http://36kr.com/p/5125134.html.

[8] "首届全国高校数据驱动创新研究大赛"评审结果公布[EB/OL]. [2018 - 05 - 01]. http://pkunews.pku.edu.cn/xwzh/2018-03/23/content_301531.htm.

[9] 国务院办公厅关于印发科学数据管理办法的通知[EB/OL]. [2018 - 05 - 01]. http://www.gov.cn/zhengce/content/2018-04/02/content_5279272.htm.

[10] Nsf data sharing policy[EB/OL]. [2018 - 10 - 01]. https://www.nsf.gov/bfa/dias/policy/dmp.jsp.

[11] 陈大庆. 英国科研资助机构的数据管理与共享政策调查及启示[J]. 图书情报工作, 2013, 57(8): 5 - 11.

[12] Science as an open enterprise[EB/OL]. [2018 - 07 - 01]. https://royalsociety.org/topics-policy/projects/science-public-enterprise/report/.

[13] 黄如花, 刘龙. 英国政府数据开放的政策法规保障及对我国的启示[J]. 图书与情报, 2017(1):1 - 9.

[14] Open science: Open access publishing and open data[EB/OL]. [2018 - 10 - 01]. http://www.aka.fi/en/funding/responsible-research/open-science/.

[15] Open access & data management[EB/OL]. [2018 - 10 - 10]. http://ec.europa.eu/research/participants/docs/h2020-funding-guide/cross-cutting-issues/open-access-dissemination_en.htm.

[16] 国务院关于印发促进大数据发展行动纲要的通知[EB/OL]. [2018 - 10 - 01]. http://www.gov.cn/zhengce/content/2015-09/05/content_10137.htm.

[17] 崔海媛, 聂华, 吴越, 等. 公共资助机构开放获取政策研究与实施——以国家自然科学基金委员会基础研究知识库开放获取政策为例[J]. 大学图书馆学报, 2017, 35(3): 79 - 86.

[18] 崔海媛, 聂华, 罗鹏程, 等. 资助机构开放获取知识库研究与构建——以国家自然科学基金基础研究知识库为例[J]. 图书情报工作, 2017,61(11): 45 - 54.

[19] Re3data. Org[EB/OL]. [2018 - 01 - 10]. https://www.re3data.org/.

[20] Datacite[EB/OL]. [2018 - 10 - 10]. https://search.datacite.org/.

[21] WHYTE A. Final results from dcc rdm 2014 survey[EB/OL]. [2018 - 08 - 01]. http://www.dcc.ac.uk/blog/rdm-2014-survey.

[22] TENOPIR C, TALJA S, HORSTMANN W, et al. Research data services in european academic research libraries[J]. Liber quar-

terly, 2017, 27(1): 23 - 44.
[23] COX A M, KENNAN M A, LYON L, et al. Developments in research data management in academic libraries: towards an understanding of research data service maturity [J]. Journal of the Association for Information Science and Technology, 2017, 68 (9): 2182 - 2200.
[24] YOON A, SCHULTZ T. Research data management services in academic libraries in the us: a content analysis of libraries' websites [J]. College & research libraries, 2017, 78(7): 920 - 933.
[25] The realities of research data management[EB/OL]. [2018 - 05 - 01]. https://www.oclc.org/research/publications/2017/oclcresearch-research-data-management.html.
[26] SERVICE T A N D. Creating a data management framework[EB/OL]. [2018 - 06 - 07]. https://www.ands.org.au/guides/creating-a-data-management-framework.
[27] CENTRE T D C. How to develop RDM services-a guide for heis [EB/OL]. [2018 - 05 - 01]. http://www.dcc.ac.uk/resources/how-guides/how-develop-rdm-services.
[28] 崔宇红. 机构研究数据管理实践探析: 模型, 核心服务和优先战略 [J]. 情报理论与实践, 2017, 40(8): 19 - 22.
[29] 殷沈琴, 张计龙, 张莹, 等. 社会科学数据管理服务平台系统选型研究——以复旦大学社会科学数据平台为例 [J]. 图书情报工作, 2013, 57(19): 92 - 96.
[30] 罗鹏程, 朱玲, 崔海媛, 等. 基于 dataverse 的北京大学开放研究数据平台建设 [J]. 图书情报工作, 2016, 60(3): 52 - 58.
[31] 刘霞, 饶艳. 高校图书馆科学数据管理与服务初探——武汉大学图书馆案例分析 [J]. 图书情报工作, 2013, 57(6): 33 - 38.
[32] WITTENBERG J, ELINGS M. Building a research data management service at the University of California, Berkeley: a tale of collaboration [J]. IFLA journal, 2017, 43(1): 89 - 97.
[33] HODSON S, MOLLOY L. Current best practice for research data management policies: a memo for the danish e-infrastructure cooperation and the danish digital library[EB/OL]. [2018 - 10 - 10]. https://ora.ox.ac.uk/objects/uuid:de360744-ad4f-458b-95cd-162eb47e6bbb.
[34] VAN ZEELAND H, RINGERSMA J. The development of a research data policy at wageningen university & research: best practices as a framework [J]. Liber quarterly, 2017, 27(1): 153 - 170.
[35] YU F, DEUBLE R, MORGAN H. Designing research data management services based on the research lifecycle-a consultative leadership approach [J]. Journal of the Australian Library and Information Association, 2017, 66(3): 287 - 298.
[36] PERRIER L, BARNES L. Developing research data management services and support for researchers: a mixed methods study [J]. Partnership, 2018, 13(1): 1 - 23.
[37] 崔海媛, 罗鹏程, 聂华, 等. 新一代学术交流生态系统的研究与构建——以北京大学为例 [J]. 图书情报工作, 2018, 63(22): 22 - 31.
[38] SAYRE F, BAKKER C, KELLY J, et al. Support for electronic lab notebooks at top american research universities [J]. Journal of eScience librarianship, 2018, 7(2): 1 - 8.
[39] 北京大学图书馆数据库导航[EB/OL]. [2018 - 10 - 20]. http://dbnav.lib.pku.edu.cn/.

作者贡献说明:
崔海媛:论文框架设计、内容撰写与论文修改;
罗鹏程:建设过程部分内容撰写与论文修改;
李国俊:研究数据管理服务主页调研部分内容撰写与论文修改;
朱玲:北京大学研究数据服务部分内容撰写。

Research and Implementation on Research Data Management Service Framework, Taking Peking University as an Instance

Cui Haiyuan Luo Pengcheng Li Guojun Zhu Ling
Peking University Library, Beijing 100871

Abstract: [**Purpose/significance**] Based on comprehensive investigation and study worldly, this paper provides the research and service framework construction of research data management. It also introduces the experience of building research data management services of Peking University. It is hoped to help academic libraries to promote research data management services for data-driven research. [**Method/process**] By analyzing the demand on data services through faculty surveys in Peking University and investigating data services in QS top 50 universities, it generalizes and designs the framework and contents of research data management services. By researching and designing the framework, it provides the framework of research data management services to academic libraires and Introduces the experiences of building the research data management Services of Peking University. [**Result/conclusion**] The paper generalizes the framework as data management plan, data services, data management, data Literacy, data long-term preservation and data policy. It is hoped to help academic libraries to promote the development of research data management service and benefit for the improvement of open access and research innovation in China.

Keywords: data driven research research data management data service framework data service data policy